Recent Updates in Two Competitive Algorithms for Model Reduction of Large-Scale Dynamical Systems

Mohammed Nizam Uddin[1]*, Mohammad Monir Uddin[2], Muhammad Hanif[1]
[1]*Department of Applied Mathematics, Noakhali Science and Technology University, Sonapur, Noakhali-3814, Bangladesh*
[2]*Department of Mathematics and Physics, North South University, Dhaka-1229, Bangladesh*

Abstract — *The modern approach, to explore scientific ideas to convince others of their validations, is through computer simulations. In simulation, one needs to convert a physical model into a mathematical model. Often also in real-life applications the mathematical models are represented by linear time-invariant (LTI) continuous-time systems. A large system leads to additional memory requirements and enormous computational efforts. Therefore,reducing the size of the system is an indispensable task for fast simulations. Various techniques are proposed in the literature to reduce the size of the large-scale LTI continuous-time systems. Among those, the Gramian based method Balanced truncation (BT) and interpolatoryprojection based method Iterative rational Krylov algorithm (IRKA) are most commonly used techniques. This article shows the comparisons between the BT andIRKA using their recent developments.*

Index Terms — *LTI continuous-time system, model reduction,balanced truncation, iterative rational Krylov algorithm.*

I. INTRODUCTION

In this article we describe generalized LTI continuous-time systems of the form:

$$Ex(t) = Ax(t) + Bu(t); \quad (1)$$

$$y(t) = Cx(t) + Du(t);$$

where the vector $x(t) \in \mathrm{R}^n$is the state vector, $u(t) \in \mathrm{R}p$is the input vector and $y(t)$ *2* $\mathrm{R}^m$is the output vector and *E.* , *BC* and *D* are all matrices with appropriatedimensions. If $p = m = 1$, the system is referred to as a single-inputsingle-output (SISO) system, otherwise it is called a multiinput multi-output (MIMO) system. In the MIMO case, we assume that the number of inputs andoutputs are much less than the number of states, i.e., $p; m << n$. Thecorresponding transfer function of this system is defined by

$$\mathrm{G}(s) = C(sE - A)^{-1}B + D, \quad (2)$$

where $s \in \mathrm{C}$. In real-life applications such mathematical models arise in different disciplines of science and engineering. See, e.g., [1], [2] for some motivating examples. In general, such systems are generated by finite element or finite difference methods (cite here). If the grid resolution becomes very fine, because many details must be resolved, the systems become very large. Moreover they are sparse, i.e., most of the elements in the matrices of the system are zero, which are not stored. A high dimensional system will always be complex, requiring a great deal of memory, thereby hindering computational performance significantly in simulation. Sometimes the systems are too large to store due to memory restrictions. Therefore, we seek to reduce the complexity of the model by applying model order reduction (MOR), i.e., we want to approximate the system (1) by the lower dimensional model:

$$\hat{E}\dot{\hat{x}}(t) = \hat{A}\hat{x}(t) + \hat{B}u(t), \hat{y}(t) = \hat{C}\hat{x}(t) + \hat{D}u(t), \quad (3)$$

where $\hat{E}$, $\hat{A} \in R^{lxl}$ ‘×‘ and $B \in \mathrm{R}^{lxp} C \in \mathrm{R}^{mxl}andD := D$ The quality of the reduced order model (ROM) relies on the difference of the two outputs *y* and *y^* which can also be measured by

$$||G(.) - \hat{G}(.)||, \quad (4)$$

where $\hat{G}(s) = \hat{C}(s\hat{E} - \hat{A})^{-1}\hat{B}, s \in C$ is the transfer functions of (3). Note that *k:k*denotes a suitable norm e.g., the H_1 or H_2 norms (see [1]). For motivations, applications, restrictions, and techniques of MOR see, e.g., [1], [2]. During the last few decades many MOR methods have been developed to find reduced order models for large-scale LTI continuous-time systems. In a broad sense, those techniques can be classified into two categories , namely,

Gramian based methods and momentmatching based methods. Recently, two prominent methods, namely the balanced truncation (BT) and the iterative rational Krylov algorithm (IRKA) based interpolatory projection method are frequently used for the model reduction of large-scale dynamical systems. Both of them have merits and demerits. Balanced truncation preserve stability of the system and it has an a-priori error bound. However, in this method one has to solve two Lyapunov equations which is computationally expensive. On the other hand, the recentlydeveloped,interpolatory method via IRKA is attractive to the model reduction community since it is computationally efficient. It requires only matrix - vector products or linear solves. Unfortunately, this prominent method has neither an a-priori error bound nor guaranteed stability preservation. In this article we show the comparisons of these two prominent techniques i.e., the BT and IRKA using their recent updates . II. BALANCED TRUNCATION (BT) A good motivation of balanced truncation can be found in [1]. The fundamental idea of balanced truncation is to truncate the *less-important states* from a system. A less-important state is a state that is difficult to control and observe. Those states essentially correspond to the smallest Hankel Singular values (HSVs) of the system. In reality, the states which are difficult to control may not be difficult to observe and vice versa. This implicates if we eliminate the states that are hard to be controlled directly from the original system, then we may also eliminate some states that are easy to observe.However,in an application, the easily observable states are essential to be preserved. The same contradiction might appear for those states that are difficult to be observed but easily controlled. This problem can be resolved by transforming the system into a balanced form. In a balanced, system the degree of controllability and the degree of observability of each state are the same. A balanced system can also be defined as follows.

Definition II.1. *A stable and minimal LTI system is called balanced if the controllability Gramian and the observability Gramian of the system are equal and diagonal. The diagonal elements are the system's HSVs.*

The system's controllability Gramin (P) and observability Gramain (Q) are the solutions of the following two Lyapunov equations

$$APE^T + EPA^T = -BB^T \quad (5)$$
$$A^T QE + E^T QA = -C^T C \quad (6)$$

respectively. Equation (5) and (6) are respectively, known as continuous-time controllability and observability Lyapunov equations. In a balanced system the relation

$$P = Q = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$$

holds true. A system can be balanced via the *balancing transformation*.

Definition II.2.A state space transformation T is called balancing transformation if it causes

$$TP\,T^* = T^{-*} Q T^{-1} = \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix}, \quad (7)$$

where P and Q are the controllability and observ -ability Gramians, $\Sigma 1 = \text{diag}(\sigma_1, \ldots, \sigma_r)$, $\Sigma_2 = \text{diag}(\sigma_{r+1}, \ldots, \sigma_n)$, *and* $\{\sigma_i\}_{i=1}^n$ *are the system's HSVs.*

Under the balancing transformations, according to theGramians in (7), the state space realization is transformed into

$$(E, A, B, C, D) \rightarrow (TET^{-1}, TAT^{-1}, TB, CT^{-1}, D) = \left(\begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}, \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, [C_1 \quad C_2], D \right).$$

Now picking up the block matrices E_{11}, A_{11}, B_1, C_1 one can form the ROM (3), where $(\hat{E}, \hat{A}, \hat{B}, \hat{C}) = (E_{11}, A_{11}, B_1, C_1)$.From the above discussion we can conclude that in the balancing based model reduction one must first compute the balancing transformation (T) to convert the system into a balanced form. Then the truncation is performed on the balanced system. For a large-scale system, balancing the whole system before truncation is infeasible. Hence, for such systems, usually the balancing and truncation are carried out simultaneously, by using the so-called *balancing and truncating* transformations. The author of [1] reviews several approaches to compute the balancing and truncating transformations among which we will focus on the *square-root method* (SRM), originally defined in [3]. To perform this method, we compute the Gramian factors $R_c$and L_c, such that $P = R_c R_c T$; $Q = L_c L T_c$ Then the balancing transformation can be formed using the SVD

$$R_c^T EL_c = U\Sigma V^T = [U_1 \quad U_2] \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

Algorithm 1: LR-SRM.

Input :E,A,B,C,D

Output:$\hat{E}, \hat{A}, \hat{B}, \hat{C}, \hat{D} := D$

1 Compute R and L as defined in (9) by solving the Lyapunov equations (5) and (6).

2 Compute SVD

$$R^T EL = U\Sigma V^T = [U_1 \quad U_2] \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

3 Construct $V := LV_1\Sigma_1^{-\frac{1}{2}}$, $W := RU_1\Sigma_1^{-\frac{1}{2}}$

4 Form

$$\hat{E} = W^T EV, \hat{A} = W^T AV, \hat{B} = W^T B, \hat{C} = CV$$

and defining

$$V := L_C V_1 \Sigma_1^{-\frac{1}{2}}, \quad W := R_c U_1 \Sigma_1^{-\frac{1}{2}}, \tag{8}$$

where U_1 and V_1 are composed of the leading k columns of U and V, respectively, $\Sigma 1$ is the first $k \times k$ block of the matrix $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k, \ldots, \sigma_n)$. Finally, by applying the balancing transformations to the system (1), one can derive the ROM (3).TheGramian factors $Rc$and $Lc$are obtained by usingthe Cholesky decompositions of the Gramians$P$ and Q. The Gramians can be computed by solving the corresponding Lyapunov equations (5) and (6). There exist direct solvers [4] as well as iterative solvers [5] to compute P and Q from (5-6).All these methods are applicable for a small dense systems. If the number of inputs and outputs are much smaller than the dimension of the system, then the Gramians$P$ and Q can usually be approximated by low-rank factors, i.e, $P \approx RR_T$and $Q \approx LL_T$(9)

where R and L are thin rectangular matrices. Therefore, instead of computing the full Gramian factors, one can compute low-rank factors of the Gramians. During the last few decades, several iterative methods were proposed, e.g., lowrank Cholesky factor - alternating direction implicit (LRCFADI) iterations [6], cyclic low-rank Smith methods [7], projection methods [8], and sign function methods [9]. Although most of the methods are shown to be applicable for large scale dynamical systems, the LRCF-ADI iteration is more attractive in the context of Gramian based model reduction for large sparse systems with few inputs and outputs. A motivation of this prominent method can be found in [10]. Readers are referred to [11, Chapter 2] to see the updated algorithm and implementing details of LRCF-ADI method.

Using the low-rank Gramian factors R and L, the squareroot method is summarized in Algorithm 1. Note that weuse [11,Algorithm 5] to compute the R and L. The reducedsystems obtained by this algorithm satisfies [1] the global error bound

$$||G - \hat{G}||_{H_\infty} \leq 2 \sum_{i=k+1}^{n} \sigma_i \tag{10}$$

where $\hat{G}$is the transfer function of the reduced model. Therelation (10) is an *a-priori* error bound. Thus, for a given errorbound (tolerance) one can use it to fix the required dimensionof the reduced system.

III. ITERATIVE RATIONAL KRYLOV ALGORITHM (IRKA)

Interpolatory projection methods seek a ROM (3) by constructing the matrices V and W in such way that the reducedtransferfunction defined in (4) interpolates the original transferfunction (2) at a predefined set of interpolation points. Thatmeans to find$\hat{G}$(s) such that

$$G(\alpha_i) = \hat{G}(\alpha_i);$$
$$C(\alpha_i E - A)^{-1}B = \hat{C}\,(\alpha_i \hat{E} - \hat{A})^{-1}\hat{B}, \text{for } i = 1, \ldots, r, \tag{11}$$

where $\alpha_i \in C$ are the interpolation points. Often, in additionto the above conditions, we are interested in matching more
quantities, that is

$G^{(j)}(\alpha_i) = \hat{G}^{(j)}(\alpha_i);$
$C[(\alpha_i E - A)^{-1}E]^j(\alpha_i E - A)^{-1}B := \quad (12)$
$\hat{C}[(\alpha_i\hat{E} - \hat{A})^{-1}\hat{E}]^j(\alpha_i\hat{E} - \hat{A})^{-1}\hat{B},$

for $j = 0,1,\ldots, q;$ where $C[(\alpha_i E - A)^{-1}E]^j(\alpha_i E - A)^{-1}B$is called the j-th moment of G(s) at α_i, and $\hat{G}^{(j)}(\alpha_i)$ represents the j-thderivative of G(s) evaluated at σ_l. Note that for$j = 0$, these conditions reduce to (11). In this paper, we restrict ourselves to simpleHermite interpolation (Cite here), where$j = 0$ and $j = 1$. In the following, we discuss how a ROM can be generated by computingfirst the appropriate projection so that a reduced interpolating approximation is guaranteed. The concept of projection forinterpolatory model reduction was first discussed in [12], and later, Grimme in [13] modified the approach by utilizing therational Krylov method [**?**]. Since Krylov based methods can achieve moments matching without explicitly computing themoments (explicit computationof moments is known to be ill-conditioned [14]), they areextremely useful for model reduction of large scale systems.
The following Lemma suggests a choice of V and W thatensure Hermite interpolation with the use of a rational Krylovsubspace.

Lemma III.1([15]). *Let us consider two sets of distinctinterpolation points,* $\{\alpha_i\}_{i=1}^r \subset C$*and* $\{\beta_i\}_{i=1}^r \subset C$,*whichare closed under conjugation (i.e., the points are either realor appear in conjugate pairs). Suppose V and W satisfy*

Range $(V) = \text{span}\{(\alpha_1 E - A)^{-1}B,\ldots,(\alpha_r E - A)^{-1}B\}$, (13a)
Range $(W) = \text{span}\{(\beta_1 E^T - A^T)^{-1}C^T,\ldots,(\beta_r E^T - A^T)^{-1}C^T\}$, (13b)
Then V and W can be chosen real and $\hat{G}(s) = \hat{C}(s\hat{E} - \hat{A})^{-1}\hat{c}$*where* $\hat{E}, \hat{A}, \hat{B}$ *and* $\hat{C}$ *are defined in (3), satisfies theinterpolation conditions*
$G(\alpha_i) = \hat{G}(\alpha_i);\ G(\beta_i) = \hat{G}(\beta_i);$ *and*
$G'(\alpha_i) = \hat{G}'(\alpha i)$ *when* $\alpha_i = \beta_i;$
for $i = 1,\ldots, r.$The subspace in (13a), that is, the span of the column vectors $(\alpha_i E - A)^{-1}B$ for $i = 1,\ldots, r$, can be considered as the union of shifted rational Krylov subspaces. For a given shift frequency $\alpha \in C$, the rational Krylovsubspace$K_q((\alpha E - A)^{-1}; (\alpha E - A)^{-1}B)$ is defined as
$K_q((\alpha E - A)^{-1}; (\alpha E - A)^{-1}B) := \text{span}\,(\alpha E - A)^{-1}B,\ldots,(\alpha E - A)^{-q}B.$
If $q = 1$ for each $(\alpha_i, i = 1,\ldots, r$, then the union of suchshifted rational Krylov subspaces is equivalent to the subspace in (13a). Analogously, the subspace in (13b) can also be defined as the union of shifted rational Krylov subspaces given above. Tosummarize, rational Krylov based model reduction requires a suitable choice of interpolation points, the construction of V and W as in Lemma III.1, and the use of Petrov-Galerkin conditions (cite here).The quality of the reduced model is highly dependent on the choice of interpolation points and therefore various techniques [12] have been developed for the selection of interpolation points. Recently in [15], the issue of selecting a good choice of interpolation points is linked to the problem of $H2$-optimal model reduction.

Definition III.2. *A ROM (3) is called H*2 *optimal if it satisfies*
$\|G\|_{H_2} = \min\|G - \hat{G}\|_{H_\infty}. \quad (14)$
$\dim(\hat{G}) = r$

Now a days IRKA becomes of of the best choice in MOR using Krylov subspace approaches [15]. Upon convergence, itidentifies a choice of interpolation points that guarantees the H_2-optimality conditions for the reduced system. Starting from an initial set of interpolation points, the IRKA iteration updates the interpolation points until they converge to a fixed value. A complete procedure of IRKA for a SISO system is given in [15, Algorithm 4.1].For model reduction of MIMO dynamical systems, rationaltangential interpolation has been developed by Gallivan et. al. [16]. The problem of rational tangentialinterpolation is to construct V and W such that the reduced transfer function$\hat{G}(s)$ tangentiallyinterpolates the original transfer functionG(s) at a predefined set of interpolation points and some fixedtangent directions. That is

$G(\alpha_i)b_i = \hat{G}(\alpha_i)b_i,\ c_i^T G(\alpha_i) = c_i^T \hat{G}(\alpha_i)$,and
$c_i^T G(\alpha_i)b_i = c_i^T \hat{G}(\alpha_i)b_i;$ for $i = 1,\ldots, r,$

where $b_i \in C^m$and $c_i \in C^p$are the right and left tangentialdirections, respectively, and both correspond to the interpolation points αi. With these quantities, the rational tangential interpolation can be achieved. The IRKA based interpolatoryprojection methods for MIMO systems have been discussed in [15], where the algorithm updates interpolation points as well as tangential directions until the reduced system satisfies the necessary condition for H_2 optimality. We have summarized a complete procedure for a MIMO system in Algorithm 2.

IV. NUMERICAL RESULTS

This section discusses the numerical results to show thecomparisons between the BT and IRKA based model reduction methods. We consider a model example; the International Space Station model (ISSM). The details of the model can be found in [17]. The dimension of the original model is 270 and the number of inputs-outputs are 3. For each model we compute different dimensional reduced models via projecting the systems onto the eigen-space of Gramians.Applying Algorithm 1 and 2 we form exemplary reduced order models of dimension 50. Figure 1 shows the frequency responses (largest singular value of $G(\omega)$ of full and the 50dimensional reduced order models obtained by BT and IRKA in frequency domain over range of 10^{-2}to10^{4}. The absolute error between the frequency responses of full and reduced

Algorithm 2: IRKA for MIMO systems.

Input :E,A,B,C,D
Output: $\hat{E}, \hat{A}, \hat{B}, \hat{C}, \hat{D} := D$.
1 Make an initial selection of $\{\alpha_i\}_{i=1}^{r}$ and $\{b_i\}_{i=1}^{r}$
$\{c_i\}_{i=1}^{r}$
2 $V = [(\alpha_1 E - A)^{-1}Bb_1, \ldots, (\alpha_r E - A)^{-1}Bb_r]$
3$W = [(\alpha_1 E^T - A^T)^{-1}C^T c_1; \cdots; (\alpha_r E^T - A)^{-1}C^T c_r.]$
4 while *(not converged)* **do**
5 $\hat{E} = W^T EV$, $\hat{A} = W^T AV$, $\hat{A} = W^T B$, $\hat{C} = CV$.
6 Compute $A\hat{z}_i = \hat{\lambda}_i E\hat{z}_i$and $y_i^* \hat{A} = \hat{\lambda}_i y_i^* \hat{E}$.
7 $\alpha_i \leftarrow \lambda_i, b_i^* \leftarrow y_i^* \hat{B}$and $c_i \leftarrow \hat{C} z_i$, for
$i = 1, \ldots, r$.
8 $V = [(\alpha_1 E - A)^{-1}Bb_1; \cdots; (\alpha_r E - A^{-1}Bb_r],$
$W = (\alpha_1 E^T - A^T)\text{-}1C^T c_1, \ldots, (\alpha_r E^T - A^T)^{-1}Cc_r.$
9 $i = i + 1$

10 $\hat{E} = W^T EV, \hat{A} = W^T AV, \hat{B} = W^T B and \hat{C} = CV.$

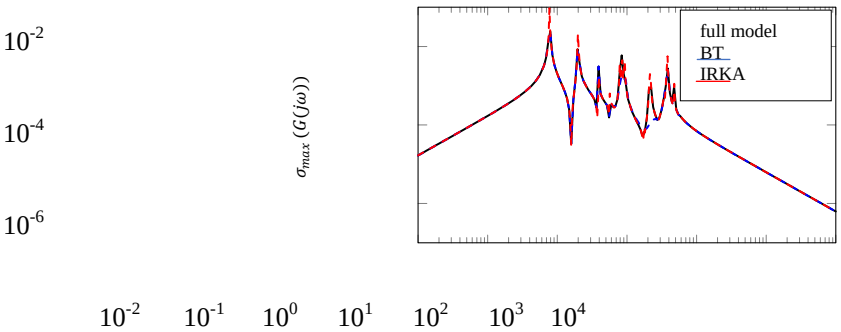

Fig. 1: Sigma plot (maximum singular values) of full andreducedorder models.

models are shown in Figure 2. In both the cases the computed errors are satisfactory. This figure deficits that in the lower frequencies the performances of the BT method is better than the IRKA. On the other hand in the higher frequencies the result is vice versa.

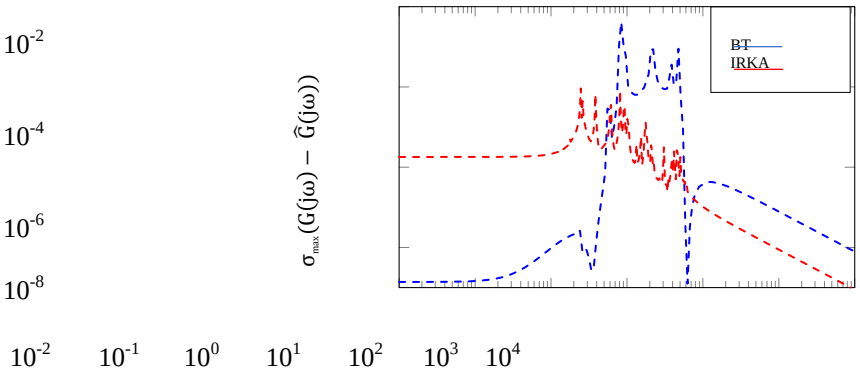

Fig. 2: Absolute error in the sigma plot of full and reducedorder models.

V. CONCLUSIONS

In this article we have discussed two prominent methods,such as balanced truncation and IRKA for model reductionof large-scale LTI continuous-time systems. The advantages and disadvantages of both methods have been discussed elaborately. For theboth methods we havepresented the updated algorithms of the MIMO systems. The comparisons are also shown by using numerical results.

REFERENCES

[1] A. Antoulas, *Approximation of Large-Scale DynamicalSystems*, ser.Advances in Design and Control. Philadelphia, PA: SIAM Publications,2005, vol. 6.

[2] P. Benner, V. Mehrmann, and D. C. Sorensen, *Dimension Reduction ofLarge-Scale Systems*, ser. Lect. Notes Comput. Sci. EngSpringer-Verlag, Berlin/Heidelberg, Germany, 2005, vol. 45

[3] M. S. Tombs and I. Postlethwaite, "Truncated balanced realization of a stable non-minimal state-space system," *Internat. J. Control*, vol. 46,no. 4, pp. 1319–1330, 1987.

[4] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation *AX+XB = C*: Algorithm 432," *Comm. ACM*, vol. 15, pp. 820–826, 1972.

[5] P. Benner and E. S. Quintana-Ort´ı, "Solving stable generalized Lyapunovequations with the matrix sign function," *Numer. Algorithms*, vol. 20,no. 1, pp. 75–100, 1999.

[6] J.-R. Li and J. White, "Low rank solution of Lyapunov equations," *SIAMJ. Matrix Anal. Appl.*, vol. 24, no. 1, pp. 260–280, 2002.

[7] T. Penzl, "A cyclic low rank Smith method for large sparse Lyapunovequations," *SIAM J. Sci. Comput.*, vol. 21, no. 4, pp. 1401–1418, 2000.

[8] Y. Saad, "Numerical solution of large Lyapunov equation," in *SignalProcessing, Scattering, Operator Theory and Numerical Methods*, M. A.Kaashoek, J. H. van Schuppen, and A. C. M. Ran, Eds. Birkhauser, ¨1990, pp. 503–511.

[9] U. Baur, "Control-Oriented Model Reduction for Parabolic Systems," Ph.D. Thesis, TechnischeUniversitat Berlin, Berlin, Jan. 2008, iSBN 978-3639074178 Vdm Verlag Dr. Muller, available from http://www.nbn-resolving.de/urn:nbn:de:kobv:83-opus-17608.

[10] P. Benner and J. Saak, "Numerical solution of large and sparse continuous time algebraic matrix Riccati and Lyapunov equations: a state ofthe art survey," *GAMM Mitteilungen*, vol. 36, no. 1, pp. 32–52, August2013.

[11] M. M. Uddin, "Computational methods for model reduction of large-scale sparse structured descriptor systems," Ph.D. Thesis,Otto-von-Guericke-Universitat, Magdeburg, Germany, 2015. [Online]. Available: http://nbn-resolving.de/urn:nbn:de:gbv:ma9:16535

[12] D. C. Villemagne and R. E. Skelton, "Model reduction using a projectionformulation," *Internat. J. Control*, vol. 46, pp. 2141–2169, 1987.

[13] E. J. Grimme, "Krylov projection methods for model reduction," Ph.D.Thesis, Univ. of Illinois at Urbana-Champaign, USA, 1997.

[14] P. Feldmann and R. W. Freund, "Efficient linear circuit analysis by Pade´ approximation via the Lanczos process," *IEEE Trans. Comput.-AidedDesign Integr. Circuits Syst.*, vol. 14, pp. 639–649, 1995.

[15] S. Gugercin, A. C. Antoulas, and C. A. Beattie, "H_2 model reductionfor large-scale dynamical systems," *SIAM J. Matrix Anal. Appl.*, vol. 30,no. 2, pp. 609–638, 2008.

[16] K. Gallivan, A. Vandendorpe, and P. Van Dooren, "Model reductionof MIMO systems via tangential interpolation," *SIAM J. Matrix Anal.Appl.*, vol. 26, no. 2, pp. 328–349, 2004.

[17] Y. Chahlaoui and P. Van Dooren, "A collection of benchmark examples for model reduction of linear time invariant dynamical systems,"University of Manchester, SLICOT Working Note 2002–2, Feb. 2002,ava